

Tratamiento de textos para mejorar la comprensión lectora en alumnos con deficiencias auditivas

Handling text in order to improve reading comprehension for hearing-impaired students

Estela Saquete, Sonia Vázquez, Elena Lloret, Fernando

Llopis, Jose Manuel Gómez, Alejandro Mosquera

Departamento de Lenguajes y Sistemas Informáticos

Universidad de Alicante

{stela,svazquez,elloret,llopis,jmgomez,amosquera}@dlsi.ua.es

Resumen: Proyecto emergente centrado en el tratamiento de textos educativos en castellano con la finalidad de reducir las barreras lingüísticas que dificultan la comprensión lectora a personas con deficiencias auditivas, o incluso a personas aprendiendo una lengua distinta a su lengua materna. Se describe la metodología aplicada para resolver los distintos problemas relacionados con el objetivo a conseguir, la hipótesis de trabajo y las tareas y los objetivos parciales alcanzados.

Palabras clave: tecnologías del lenguaje humano, comprensión lectora, simplificación textual

Abstract: This project is focused on textual treatment in Spanish in order to reduce language barriers that hinder hearing impaired people from reading comprehension, or even people learning a new language. This paper describes the methodology used to face the different problems related to the proposed objective, as well as the working hypothesis and partial tasks and objectives achieved.

Keywords: human language technologies, Reading comprehension, textual simplification

1 Datos del proyecto

Este proyecto está dirigido por Estela Saquete, miembro del Grupo de Procesamiento del Lenguaje y Sistemas de Información (GPLSI) de la Universidad de Alicante. Está financiado por la Universidad de Alicante (GRE11-21) dentro del programa de ayudas a proyectos emergentes.

Contacto

Email: stela@dlsi.ua.es

Teléfono: 965903400 ext. 2894

Dpto. de Lenguajes y Sistemas Informáticos,
Universidad de Alicante,

Carretera San Vicente del Raspeig, s/n,
03690, Alicante, España

2 Introducción

La comprensión lectora se define en el informe de PISA 2000¹ como: *"la capacidad de comprender, emplear información y reflexionar a partir de textos escritos, con el fin de lograr las metas individuales y desarrollar el conocimiento y el potencial personal y así participar eficazmente en sociedad"*.

Es por ello que, actualmente, dicha comprensión lectora se ha convertido en uno de los principales temas de estudio en el ámbito de la psicología y la educación (Olivé, 2009). Dichas investigaciones han determinado que las habilidades y condiciones que se requieren para la comprensión lectora son complejas y múltiples, siendo fundamental dentro de este

¹ <http://www.mec.es/multimedia/00005713.pdf>

grupo las habilidades lingüísticas, y que son en las que nos centraremos para este proyecto.

La escasa capacidad de comprensión lectora es un problema que va en aumento en nuestra sociedad, como ya se constató en el último informe PISA (Programa para la Evaluación Internacional de Alumnos) 2006. Además, el problema de la comprensión lectora todavía es mayor para las personas con deficiencias auditivas. Esta problemática que relaciona las dificultades en la comprensión lectora con la sordera lleva tiempo siendo estudiada, tanto a nivel léxico como sintáctico (King & S.P., 1985) (LaSasso & Davey, 1987) (Paul & Gustafson, 1991) (Berent, 1996).

Estudios previos han detectados los siguientes obstáculos que las personas con deficiencias auditivas se encuentran a la hora de comprender un texto (Mies, 1992) (Herrera, 2003) (Stockseth, 2002):

a) *problemas de ambigüedad*. Existen numerosas palabras en castellano que pueden poseer más de un significado, y toman su sentido correcto en función del contexto en el que aparecen. Todo este tipo de palabras polisémicas pueden generar muchos problemas para la comprensión de un texto;

b) *vocabulario limitado*. Estos lectores se fijan más en palabras familiares y usan más sustantivos concretos y verbos familiares que palabras abstractas. Muchas veces también tienen problemas con el reconocimiento de entidades de nombre y su contextualización;

c) *dificultades en la interpretación de estructuras sintácticas complejas* cuya sintaxis se desvía de la estructura básica de constituyentes explícitos sustantivo-verbo-sustantivo y del orden Sujeto-Verbo-Objeto. Es decir, estructuras más complejas, como por ejemplo, oraciones activas simples transitivas, pasivas o coordinadas o subordinadas suponen un problema de comprensión extra para personas con deficiencias auditivas;

d) *problemas en situar los eventos en la línea temporal* si en el texto existen saltos en el tiempo, hacia delante o hacia atrás, que implican una interpretación de las señales y expresiones temporales para su total comprensión.

Como suele ocurrir, muchos de estos estudios se centran en la problemática lingüística del inglés, pero afortunadamente también existen numerosos trabajos que estudian la comprensión lectora en castellano, tanto a nivel léxico (Mies, 1992) como a nivel

sintáctico (Stockseth, 2002). Además, hay múltiples estudios en castellano especialmente aplicados al ámbito educativo (Alegria & Leybaert, 1985) (Asensio & Carretero, 1989) (Ferrández Mora, 1989).

Dentro del área de investigación que nos ocupa, las Tecnologías del Lenguaje Humano, existe una amplia variedad de trabajos relacionados con la obtención de manera automática o semi-automática del lenguaje de signos a partir de textos hablados o escritos (Parton, 2005) (Wu, Chiu, & Guo, 2004) (Duchnowski, Lum, Krause, Sexton, Bratakos, & Braid, 2000). Además, debemos destacar el proyecto MÁS, financiado por la Unión Europea, y cuya finalidad es comprobar los efectos del uso de una herramienta multimedia para mejorar la comprensión lectora utilizando como apoyo el lenguaje de signos (Ferrer Manchón, 2001).

3 Objetivos del proyecto

El objetivo principal de este proyecto es diseñar, implementar y evaluar una herramienta tecnológica que sea capaz de convertir documentos en castellano en textos de lectura fácil para personas con deficiencias auditivas, y por tanto, pueda servir de ayuda para la comprensión lectora de dichas personas. El proceso de conversión implica detectar automáticamente características lingüísticas del documento de entrada que puedan impedir su fácil comprensión y automáticamente reducir y/o eliminar esos obstáculos pero preservando el significado general del documento en la medida de lo posible. Las tecnologías del lenguaje humano serán aplicadas para eliminar los obstáculos en la comprensión derivados de las estructuras complejas y de la ambigüedad en los significados en textos escritos, y generando un apoyo extra mediante imágenes o pictogramas, líneas temporales, definiciones de enciclopedias online simples y resúmenes del texto original.

Por tanto, los objetivos concretos del proyecto serán:

1. Analizar en profundidad las diferentes aproximaciones existentes en las Tecnologías del Lenguaje Humano para tratar cada uno de los obstáculos que dificultan la comprensión de los textos.

2. Desarrollar las herramientas necesarias para poder solucionar los problemas analizados y evaluar cada una de estas herramientas de manera independiente. Además, también serán evaluadas en las diferentes competiciones relacionadas con el fin de demostrar su funcionalidad y eficiencia.

3. Desarrollar una interfaz que, dado un texto de entrada, aplique las diferentes herramientas desarrolladas para obtener un texto de lectura fácil con los apoyos que hemos comentado y reduciendo y/o eliminando las posibles barreras del texto.

4. Analizar y evaluar la herramienta con los usuarios finales, con el fin de determinar su eficacia, su usabilidad y su impacto en la comprensión lectora de estos usuarios.

Tal y como hemos presentado en el apartado anterior, las herramientas tecnológicas existentes hasta el momento para facilitar la comprensión de textos a las personas con sordera se basan principalmente en el uso del lenguaje de signos (Parton, 2005) (Wu, Chiu, & Guo, 2004) (Duchnowski, Lum, Krause, Sexton, Bratakos, & Braida, 2000). Sin embargo, nuestra propuesta, pretende facilitar al lector el significado del texto, transformándolo a un texto de lectura sencilla, facilitando por un lado la comprensión léxica, y por otro la comprensión sintáctico-semántica. Para la comprensión léxica se detectarán aquellas palabras más complejas y se aportará, en dichos casos, de manera automática, un conjunto de posibles palabras sinónimas más comunes² y un pictograma relacionado con la palabra³. Además, para las entidades nombradas se facilitarán tanto imágenes relacionadas como una definición obtenida de una enciclopedia online simple⁴. Por otro lado, a nivel sintáctico-semántico, la herramienta detectaría automáticamente estructuras sintácticas complejas e intentaría traducirla a una o varias estructuras más sencillas, así como la detección y resolución de expresiones temporales y la generación de resúmenes que simplifiquen textos complejos.

Esta herramienta supone un interés muy importante para las personas con deficiencias auditivas, puesto que podría ayudarles a mejorar en su comprensión lectora y por tanto permitir a estas personas ampliar sus horizontes

informativos y culturales. Además, puede ser una ayuda muy indicada en el caso de la educación de los niños con sordera, en quienes los mismos profesores suelen realizar manualmente esta tarea de conversión para facilitar una lectura fluida con menor esfuerzo y más provecho para estos niños.

4 Tareas a desarrollar

Para la consecución del proyecto será necesario completar el conjunto de tareas que se mencionan a continuación:

Análisis del problema

En esta tarea se analizarán las distintas aproximaciones existentes a la desambiguación de significados, al tratamiento de estructuras complejas, a la recuperación de definiciones e imágenes asociadas a un significado concreto, palabras, o frases y así como la asociación de significados a pictogramas. También analizaremos el problema de la resolución de información temporal y la generación de resúmenes sencillos de fácil comprensión. Sobre esta base teórica se investigarán nuevas técnicas para la mejora del sistema, basándonos en la adquisición de conocimiento general del mundo.

Desarrollo y evaluación

En esta tarea se llevará a cabo la implementación de las técnicas estudiadas en la tarea anterior, en módulos independientes, dando como resultado un sistema capaz de detectar entidades complejas en un texto, desambiguarlas, proporcionar pictogramas asociados y la línea temporal asociada, además de generar un pequeño y simple resumen del texto de forma automática. En este punto se realizará también la evaluación de nuestra investigación en las diferentes competiciones internacionales relacionadas.

Visualización de la información

En esta tarea se busca crear un interfaz sencillo y muy amigable que permita al usuario acceder a toda esta información extra proporcionada automáticamente a partir de un texto de entrada en castellano y que pretende hacer más sencillo su entendimiento. En concreto, esta interfaz será puesta a disposición del Centro de Apoyo al Estudiante (CAE) de la

² <http://es.wiktionary.org>

³ Google imágenes

⁴ <http://simple.wikipedia.org>

Universidad de Alicante para que pruebe la herramienta con este colectivo de la Universidad, realizando de esta forma una evaluación de la misma, que medirá su eficiencia, usabilidad y eficacia en el soporte a la comprensión de textos, recibiendo así mismo una retroalimentación para su mejora.

5 Situación actual del proyecto

Dentro de las tareas antes mencionadas, actualmente se ha realizado parte del desarrollo del sistema propuesto.

Hasta el momento, el sistema carga el texto en castellano y es capaz de reconocer tanto las entidades nombradas como las expresiones temporales, ambos elementos del texto son marcados con un enlace. Si el usuario final tiene problemas en el entendimiento de la entidad nombrada o de la expresión temporal elegiría el término y obtendría lo siguiente:

- *Para las entidades nombradas:* una definición del término en Simple Wikipedia y tres imágenes de Google imágenes.
- *Para las expresiones temporales:* la expresión temporal resuelta en una fecha o periodo de fechas concretas.

El sistema será completado con los avances que se vayan realizando para completar todas las tareas propuestas para el proyecto.

Bibliografía

- Alegría, J., & Leybaert, J. (1985). Adquisición de la lectura en el niño sordo: un enfoque psicolingüístico. *Investigación y Logopedia*.
- Asensio, M., & Carretero, M. (1989). La lectura en los niños sordos. *Cuadernos de pedagogía* 174.
- Berent, G. (1996). The acquisition of English Syntax by Deaf Learners. *Handbook of Second Language Acquisition*, 469-506.
- Duchnowski, P., Lum, D., Krause, J., Sexton, M., Bratakos, M., & Braid, L. (2000). Development of speechreading supplements based on automatic speech recognition. *IEEE Trans Biomed Eng.* 47(4), 487-496.
- Ferrández Mora, J. A. (1989). La lectura en el currículum escolar del niño sordo. *Rev. Logop Fon Audiol*, IX.
- Ferrer Manchón, A. (2001). La comprensión lectora en personas sordas adultas y el acceso a la Universidad. ISAAC 2001: Odissea de la Comunicación. Segundas Jornadas sobre comunicación Aumentativa y Alternativa.
- Herrera, V. (2003). Desarrollo de habilidades lectoras en sujetos sordos signates, a partir del uso de código dactílicos. Tesis doctoral no publicada.
- King, C., & S.P., Q. (1985). *Reading and Deafness*. College Hill Press.
- LaSasso, C., & Davey, B. (1987). The relationship between lexical knowledge and reading comprehension for prelingually, profoundly hearing impaired students. *The Volta Review* 89, 211-220.
- Mies, B. (1992). El léxico en la comprensión lectora: Estudio de un grupo de alumnos sordos adolescentes. *Rev. Logop., Fon., Audiol.*, Vol. XII.
- Olivé, M. J. (2009). La lectura: una necesidad para la inclusión social y la democracia. Separata del Manifiesto PIAPAS. Confederación española de familias de personas sordas.
- Parton, B. (2005). Sign language recognition and translation: a multidisciplinary approach from the field of the artificial intelligence. *J. Deaf Stud Deaf Educ.*
- Paul, P., & Gustafson, G. (1991). Hearing-impaired students' comprehension of high-frequency multi-meaning words. *Remedial and special education*, 12, 52-62.
- Stockseth, D. R. (2002). Comprensión de la sintaxis española por lectores sordos chilenos. *Revista Signos*, v. 35.
- Wu, C., Chiu, Y., & Guo, C. (2004). Text generation from Taiwanese Sign Language using PST-based language model for augmentative communication. *IEEE Trans Neural Syst. Rehabil Eng.*, 12(4).